산업용 음성 DB를 위한 XML 기반 메타데이터*

주영희(성신여대), 홍기형(성신여대)

<차 례>

1. 서론
2. 산업용 음성 DB 모델
3. 산업용 음성 DB 메타데이터 모델
4. XML 기반 음성 DB 메타데이터
5. 결론 및 향후 과제

<Abstract>

XML Based Meta-data Specification for Industrial Speech Databases

Young-Hee Joo, Ki-Hyung Hong

In this paper, we propose an XML based meta-data specification for industrial speech databases. Building speech databases is very time-consuming and expensive. Recently, by the government supports, huge amount of speech corpus has been collected as speech databases. However, the formats and meta-data for speech databases are different depending on the constructing institutions. In order to advance the reusability and portability of speech databases, a standard representation scheme should be adopted by all speech database construction institutions. ETRI proposed a XML based annotation scheme [5] for speech databases, but the scheme has too simple and flat modeling structure, and may cause duplicated information. In order to overcome such disadvantages in this previous scheme, we first define the speech database more formally and then identify objects appearing in speech databases. We then design the data model for speech databases in an object-oriented way. Based on the designed data model, we develop the meta-data specification for industrial speech databases.

* Keywords: Speech DB, Standardization, Metadata, XML.

* 본 연구는 한국SIT산업협회의 "표준 산업용 음성DB 메타데이터 및 구축도구 개발" 사업으로 수행되었다.

1. 서 론

산업용 음성 데이터베이스(이하 산업용 음성 DB)는 자동차, 로봇, 가전제품 등 다양한 산업 제품을 위한 음성 사용자 인터페이스(음성인식/합성)의 개발에 있어서 필수적이며 가장 첫 단계에서 구축되어야 한다.

다양한 산업 제품의 사용 환경을 반영한 산업용 음성 DB의 효과적인 구축 및 활용은 음성 사용자 인터페이스의 효율적인 개발과 성능향상에 있어 매우 중요하며, 다양한 산업 제품의 사용자가 필요로 하는 음성언어정보가 충실히 표현된 DB, 개발자가 기술개발에 용이하게 활용할 수 있는 구조화된 DB 구성이 요구된다. 그러나 지금까지 국내에서는 산업용 음성 DB 표준화에 대한 필요성은 충분히 인지하고 있었으나 구체적인 노력이 미미하였다. 최근 대량의 산업용 음성 DB 구축이 진행되고 있으나, 각기 다른 기관에서 구축한 DB의 호환성에 문제가 발생하고 있다. 이에 따라 산업용 음성 DB의 호환성을 높이고, 활용성을 증대시키기 위해 메타데이터 표기방법 및 구조의 표준화가 절실히 필요한 시점이다.

각 업체, 연구소, 대학에서 구축한 산업용 음성 DB의 표기를 통일함으로써 DB 유통 활성화를 꾀할 수 있어 국가적으로 자원의 효율적 활용이 가능하다. 구체적인 필요성은 다음과 같다.

첫째, 산업용 제품의 음성인식/합성 엔진 개발자들이 용이하게 다양한 음성 DB를 쉽게 활용할 수 있으므로, DB 사용의 편의성이 증대되어 DB 전처리에 소요되는 시간과 인력을 줄일 수 있다.

둘째, 각 업체, 연구소, 대학에서 개별적으로 구축한 산업용 음성 DB의 경우, 현재까지는 개별 기관별로 독자적인 정보 표현 양식을 사용함으로써, 유사한 DB의 중복 구축이 많으며, 상호 호환되지 않음으로써 공유하기에는 별도의 변환 프로그램을 개발하여야 하는 등의 문제가 있으나, 본 통일된 메타데이터의 사용을 통하여 이러한 문제를 해결할 수 있다.

셋째, 음성 DB의 경우에 그 크기가 클수록 음성처리 기술의 성능을 향상시키는 데 도움이 되므로, 표준 메타데이터를 통하여, 기 구축한 유사 DB를 통합하여 대량의 DB 구축이 가능하다면, 음성처리 기술의 성능 향상에 크게 기여할 것이다.

본 논문은 산업용 음성 DB 메타데이터 표준안으로, 산업용 음성 DB의 구축 환경 정보, 구축에 참여한 화자(speaker) 정보, 화자가 발성한 단어 또는 문장 목록, 음성의 디지타이징 방법 및 저장 음성 파일의 개별 정보, 그리고 다양한 전사 정보의 표준 표시 방법을 제시한다. 세계적 공개 표준인 XML(eXtended Markup Language) 기반으로 개발하여, 향후 확장성을 보장하고, 표준화된 메타데이터 관리 및 구축 도구의 개발을 통하여 표준의 활성화를 꾀하고자 한다.

국외의 경우 프랑스의 ELRA(The European Language Resources Association)[1]와 미국의 LDC(The Linguistic Data Consortium)[2]를 주축으로 음성 DB 구축 및 표준

화를 꾸준히 수행해 오고 있다. 표준화 대상으로는 수집시스템, 수집환경, 음성DB, 전사방법, 검증방법 등 다양한 분야에서 수행하고 있다. 그러나 국제 표준화 활동은 영어권의 음성 특성을 반영하고 있어, 우리말 음성의 고유한 특성을 반영해야 할 국내 표준화에 그대로 채택하여 사용하기에 적합하지 않은 부분이 있다[3].

국내의 경우 음성 DB 메타데이터 표준안은 2003년도에 ETRI(한국전자통신연구원) 음성/언어정보 연구센터에서 개발한 공통 음성 DB 표준안[4][5]이 있으나, 음성 DB 표준안은 정보통신망 기반 음성 기술 개발을 위한 음성 DB 메타데이터 규격으로 개발되었다. 음성 DB 정보를 크게 기본정보, 음성정보, 전사정보, 화자정보, 환경정보, 파일정보, 기타정보로 나누었다. 공통 음성 DB 표준안은 국내에서 처음으로 시도한 음성 DB 메터데이터 규격이라는 점에서 의미가 있으나, 다음과 같은 단점이 있다.

첫째, 음성 DB 메타데이터를 단순히 2단계 구조로 정의하였다. 이러한 평면적인 구조는 음성 DB를 구성하는 객체 사이의 집합적 포함관계나 연관관계를 정확하게 표현하기에 부족하다. [5]에서 기술한 모델은 하나의 음성 파일에 대한 것인지, 음성 DB, 즉 다수의 음성 파일에 대한 것인지가 불명확하다.

둘째, 화자, 음성 녹음채널, 음성 파일 등 DB를 구성하는 객체의 식별자와 이를 참조하는 객체 참조 개념이 결여되어 있어, 동일한 정보의 중복 및 특정 음성 파일의 화자의 식별이 어렵다. 예를 들어, 발화자가 100명이고 발화 문장이 100개가 존재할 경우 메타데이터에 총 10,000개의 음성 파일에 대한 정보가 존재하게 된다. 이 경우 발화자 한 명당 발화 문장 100개씩을 메타데이터 문서에 중복하여 입력해야 한다.

본 논문에서 제안하는 산업용 음성 DB 메타데이터는 국제기관인 LDC, ELRA의 용어 및 분류정보를 참조하고 [5]의 모델을 보다 객체 지향 구조로 보완하였다.

본 논문의 구성은 다음과 같다. 2장에서 산업용 음성 DB의 메타데이터 필요성을 기술하고 3장에서 제안된 산업용 음성 DB 모델을 제시한다. 4장에서 산업용 음성 DB 메타데이터 모델링을 기술하며 5장에서 XML 스키마를 사용하여 정의한 메타데이터를 설명하고 그에 적합한 인스턴스 문서 예제를 제시한다. 마지막으로 6장에서 결론 및 향후 계획을 기술한다.

2. 산업용 음성 DB 모델

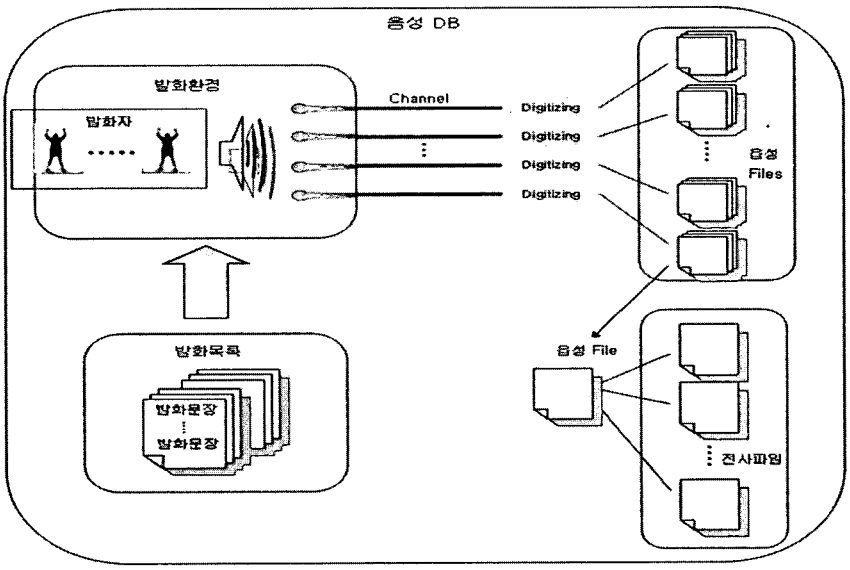
자동차, 완구, 가전 기기, 산업용 기기 등 다양한 소음 및 주변 환경에서 높은 성능의 음성 기술(인식/합성/인증/코딩/강화) 개발 및 시험을 위해서는, 해당 산업용 음성 기술이 사용되는 동일한 환경에서 대량의 음성 데이터베이스의 구축이 필수적이다. 즉, 자동차에서 사용하려고 하는 음성 인식 기술의 개발을 위해서는

자동차의 주행 환경에서 사용자가 내릴 수 있는 음성 명령을 녹음하여, 기술의 개발에 사용하여야 자동차 환경에서 성능이 높은 인식 기술을 개발할 수 있다. 또한 음성은 화자(speaker)의 특성에 따라 동일한 환경이라고 하더라도 많은 차이가 있다. 사투리의 구사여부, 개인의 구강 구조의 특성, 성별, 나이 등에 따라 동일한 환경, 동일한 단어를 발성하더라도 음성 신호에서 많은 차이를 보이므로, 화자 독립형의 음성 기술의 개발에서는 가능한 많은 수의 화자로부터 음성을 수집하여야 한다. 따라서 완구용 음성 기술을 위한 음성 DB, 자동차용 음성 기술을 위한 음성 DB 등과 같이 산업용 음성 기술 개발을 위해서는 해당 기기의 사용 환경에서 많은 수의 화자로부터 음성을 수집하는 것이 필수적이다.

산업용 음성 **DB**는 산업 응용에서 적합한 음성 기술(인식/합성/인증/코딩/강화) 개발 및 시험을 위하여, 다수의 화자로부터 채취한 음성 파일의 집합으로 정의한다.

산업용 음성 **DB** 메타데이터는 산업용 음성 DB의 기본정보 (구축 기관정보 등), 음성파일의 디지타이징 방법, 화자 정보 등과 같이 산업용 음성 DB에 대한 정보를 말한다.

본 메타데이터 규격의 대상이 되는 산업용 음성 DB는 <그림 1>과 같은 구성을 전제로 한다.



<그림 1> 산업용 음성 DB 메타데이터 구성

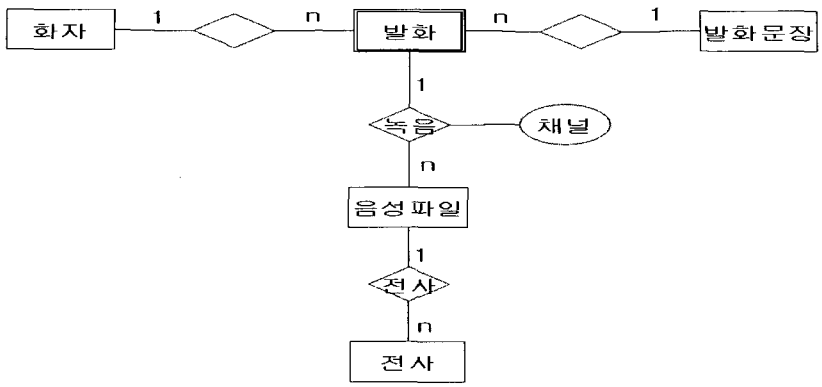
1) 음성 DB 내의 음성 파일은 다양한 발화환경에서 발화한 것을 수집한 것으로 가정하였다. 발화환경은 방음실, 백화점, 자동차 안, 가정과 같은 화자의 주변 환경을 의미한다.

2) 하나의 음성 DB는 화자의 음성을 여러 가지 방법으로 수집하기 위해서 다수의 녹음 채널이 있을 수 있다. 채널이란 녹음 시 사용하는 마이크, 디지타이징 형식, 통신망(전화망, IP network 등), 발화환경 등을 서로 다르게 할 수 있다.
3) 하나의 음성파일은, 화자가 발화목록의 한 문장을 발화한 음성을 특정한 채널로 녹음하고 디지타이징하여 저장한 것이다.
4) 하나의 발화목록은 여러 개의 발화문장으로 구성될 수 있다.
5) 하나의 음성 파일은 여러 종류의 전사정보를 가질 수 있다.
6) 화자 한 명은 다수의 발화 목록을 발성할 수 있다.
7) 발화시점을 다르게 하여 화자는 같은 발화목록을 여러 번 반복 발성하여 녹음할 수 있다.

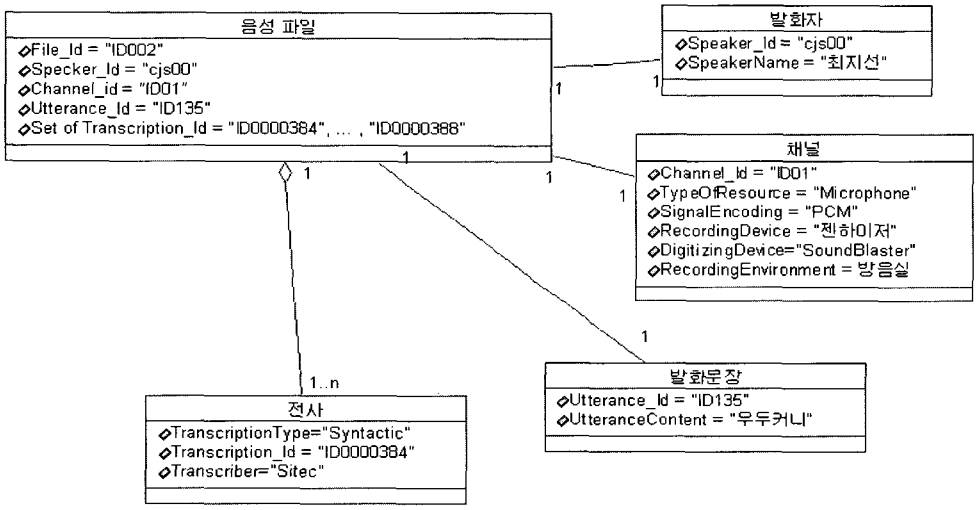
3. 산업용 음성 DB 메타데이터 모델

음성 DB의 구성요소는 화자, 발화문장, 채널, 음성파일, 전사정보이다. 음성 DB 메타데이터 구성요소들의 관계는 <그림 2>와 같다. 화자 한 명이 여러 개의 문장을 반복 발화할 수 있으므로 발화자와 발화는 1:n의 관계이다. 발화문장 하나는 여러 화자가 반복하여 발화하므로, 발화와 발화문장의 관계는 n:1이 된다. <그림 2>에서 '발화'는 화자가 발화하는 행위 자체를 하나의 개체로 바라보기 위한 약한 개체(weak entity)이다. 화자 한 명이 하나의 발화문장을 발화할 때, 다수의 채널로 녹음할 수 있으므로 채널별로 음성파일이 생성되어 발화와 음성파일은 1:n의 관계이다. 이렇게 생성된 한 개의 음성파일에는 전사정보가 전사의 종류에 따라 여러 개 존재할 수 있다. 따라서, 하나의 음성파일과 전사정보는 1:n의 관계를 가진다.

하나의 음성파일과 관련 있는 화자, 채널, 발화문장, 전사정보와의 구체적인 관계는 <그림 3>과 같다. 음성파일에는 식별자(*File_Id*)가 있으며, 누가 발화하였는지 나타내기 위해 발화자의 식별자(*Speaker_Id*)를 참조한다. 어느 채널을 통해 녹음된 파일인지 나타내기 위해 채널의 식별자(*Channel_Id*)를 참조한다. 어떤 문장에 대한 발화인지를 나타내기 위해 발화문장 식별자(*Utterance_Id*)를 참조하며, 해당 음성파일과 관련한 전사정보를 모두 참조(Set of *Transcription_Id*)하고 있다.



<그림 2> 음성DB 메타데이터 구성요소 및 관계도 (ER 모델)



<그림 3> 음성 파일 하나와 다른 정보와의 관계도

4. XML 기반 음성 DB 메타데이터

3장에서 살펴본 바와 같이 산업용 음성 DB는 수집채널, 화자 정보, 발화목록, 음성파일, 전사 정보로 구성된다. 제안하는 메타데이터 규격은 기존의 SiTEC과 ETRI에서 추진한 표준화 작업[4][5]과 비교하여 XML[6]의 특징을 최대한 살릴 수 있도록 정의하였다. 화자, 채널, 발화목록, 파일, 전사 정보에 각각의 식별자를 두고 구성요소 사이의 관계를 식별자의 참조로 표시할 수 있도록 하였다. 또한 연관되는 엘리먼트들의 그룹핑과 계층구조를 도입하여, 메타데이터의 구조를 체계화하여, 향후 음성 DB의 검색이나 재사용이 용이하도록 설계하였다.

XML 스키마로 정의한 메타데이터 구조는 <그림 4>와 같다. 본 메타데이터 정의에서 사용한 엘리먼트와 속성의 이름은 ELRA와 LDC에서 부분적으로 참고하였다[1][2]. <그림 4>의 ①번, SpeechDBMetadata는 메타데이터인 XML 문서의 루트 엘리먼트이다. ②번 General은 해당 음성 DB의 기본정보를 위한 엘리먼트이다. ③번 Channels는 음성 DB 수집에서 사용한 다수의 채널에 대한 정보를 위한 엘리먼트이다. Channels 엘리먼트의 하위 엘리먼트로 Channel 엘리먼트를 다수 가질 수 있으며, 각 Channel 엘리먼트는 특정 수집채널에 대한 정보를 기술한다. ④번 Speakers는 화자 정보를 나타내는 엘리먼트이다. Speakers는 다수의 Speaker 엘리먼트를 가질 수 있고, 하나의 Speaker 엘리먼트는 특정 화자의 정보에 해당한다. ⑤번 UtteranceLists는 DB 수집에서 사용한 발화목록 정보이다. 하나의 DB 수집에서 다수의 발화목록이 있을 수 있으므로, 하위 엘리먼트로 다수의 UtteranceList 엘리먼트를 가진다. ⑥번 Files는 DB를 구성하는 각 음성파일에 대한 정보를 담고 있는 File 엘리먼트를 DB에 존재하는 음성파일의 개수만큼 하위 엘리먼트로 가진다. ⑦번 Transcription은 상위 엘리먼트인 File이 기술하고 있는 음성파일에 존재하는 전사정보를 위한 엘리먼트이다. 하나의 음성파일은 하나 이상의 전사정보가 있을 수 있으므로, 하나의 File 엘리먼트는 다수의 Transcription 엘리먼트를 가질 수 있다. <그림 3>에서 Set of *Transcription_Id*는 XML 스키마로 구현하면서 File의 하위 엘리먼트인 Transcription을 정의하는 구조로 표현하였다.

4.1. 기본정보 (General 엘리먼트)

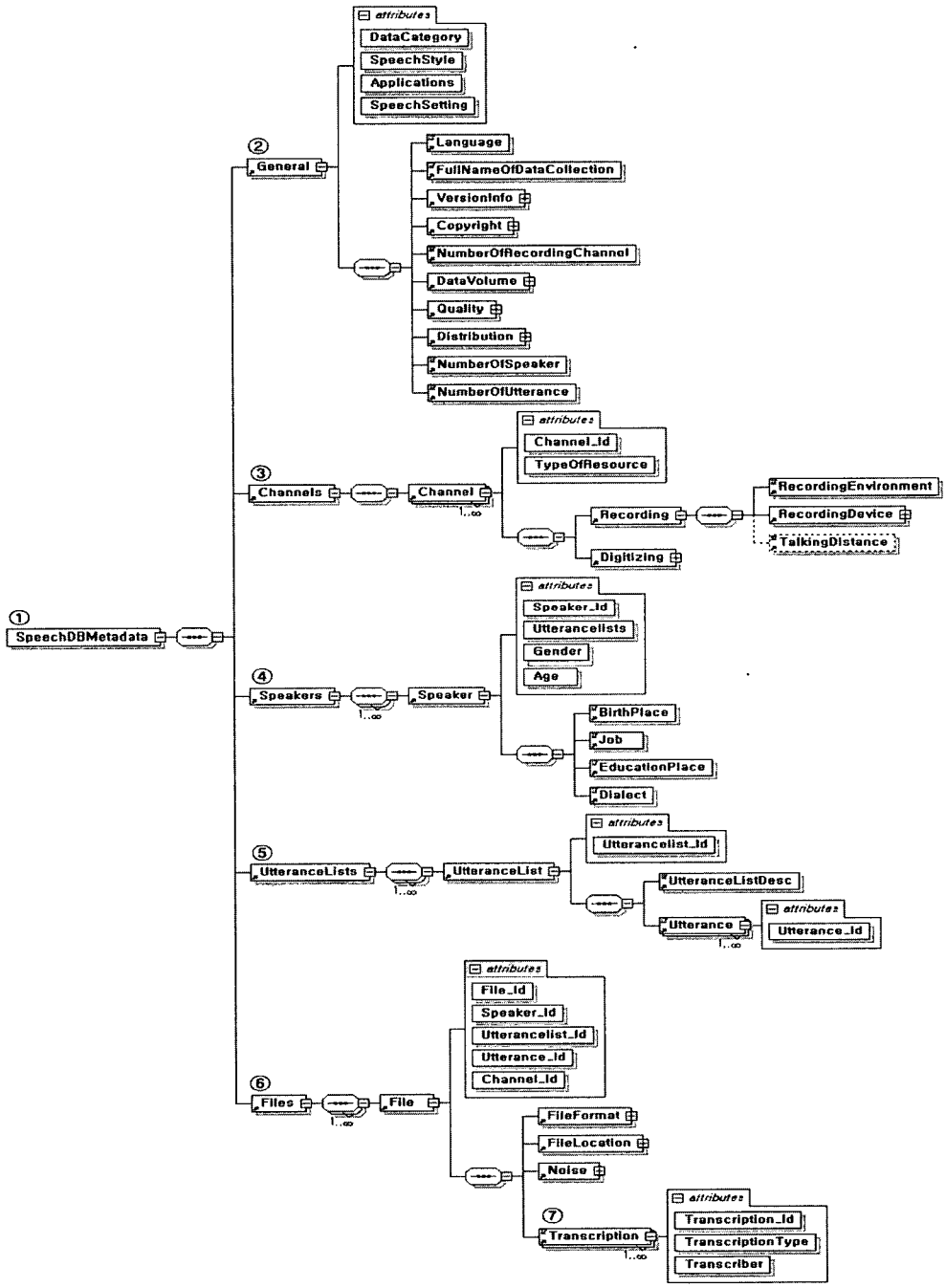
기본 정보는 전체 음성 DB에 해당하는 공통 정보, 저작권에 관한 정보, 배포할 시 필요한 정보 및 발화자 정보나 발화목록 정보로부터 유도될 수 있는 정보로 나누어질 수 있다.

<표 1>은 기본정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다. 표에서 속성은 이탤릭체로 표기하였고 엘리먼트는 정자로 표기하였다.

4.2. 수집채널 (Channel 엘리먼트)

수집채널은 화자가 발성하는 음성을 어떤 방식으로 조작하고, 저장하였는지를 의미한다. 하나의 음성 DB는 여러 개의 채널로부터 구축될 수 있다. 따라서 다채널 설정이 가능하고, 각 채널 별로 레코딩 방식과 디지타이징 방식을 다르게 할 수 있다. 각 채널은 녹음 시 발화환경을 달리하여 설정할 수 있다. 발화환경은 녹음환경이 사무실, 야외, 백화점 등과 같은 어떤 장소인지를 명시한다. *Channel_Id*는

채널을 구별할 수 있는 식별자이다. 수집채널은 크게 디지타이징 방법과 레코딩 방법으로 분류하여 정의하였다. <표 2>는 수집채널의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.



<그림 4> 산업용 음성 DB 메타데이터 스키마 구조

<표 1> 기본정보 엘리먼트 및 속성 (엘리먼트 : 정자, 속성 : 이탤릭체)

분류	엘리먼트 및 속성	내용
음성 DB 공통 정보	Language	음성DB 발성에 사용된 언어를 기록하는 것으로 ISO 3166 표기 방식
	FullNameOfDataCollection	음성 DB의 이름
	VersionInfo	버전과 음성DB 버전별로 녹음 구축기간
	NumberOfRecordingChannel	동시녹음 채널 수
	DataCategory	음성 DB의 분류 (가능한 값 : IsolatedWord,Digit, ReadSpeech, PsuedoReadSpeech, Interaction, Text, Verification, SpeechSynthesis, SpeakerCircumstance, Other, Orthography, Accent, FreeAnswerQuestion, Syllables, VCVSequence, IsolatedDigits, Continous Sentences, PRS, PBS, YesNoQuestions)
	SpeechStyle	발화할 내용의 스타일, 종류 (가능한 값 : Spontaneous, Read, Elicited, Prepared, Prompted, Other)
	Applications	상용서비스를 고려할 때 본 음성DB의 응용분야 (가능한 값 : DiscourseAnalysis, VoiceControl, LanguageIdentification, SpeakerIdentification, SpeakerVerification, SpeechRecognition, SpokenDialogue Systems, Other)
	SpeechSetting	단독발화(monologue), 대화체(dialogue), 3인이상 대화체(multilogue) 구별
저작권	Authors	음성DB 저작자의 이름
	Distributor	음성DB 설계, 구축을 주관한 기관명
	Performer	음성DB를 최종 배포한 기관명
배포시 필요한 정보	DataVolume	DB전체 크기 및 재생시간
	Distribution	DB압축 여부, 배포방법, 배포분류
	Quality	품질상태, SNR, ClippingRate와 같은 음성 DB 품질
유도 정보	NumberOfSpeaker	발성한 총 발화자 수
	NumberOfUtterance	총 화자가 발성한 총 발화 횟수

4.3. 화자 (Speaker 엘리먼트) 정보

화자는 여러 명이 존재할 수 있고 각각의 화자는 *Speaker_Id*로 구분한다. 발화자 한 명은 여러 개 발화목록을 발화할 수 있다. 특정 화자가 발화한 목록은 *Utterancelists* 속성으로 나타낸다. *Utteranclists* 속성은 발화목록의 식별자인 *Utterancelist_Id*를 집합 참조한다. <표 3>은 발화자 정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

<표 2> 수집채널 엘리먼트 및 속성

분류	엘리먼트 및 속성	내용
채널 정보	*Channel_Id*	채널 식별자
	TypeOfResource	수집환경 중 유선/무선전화망, VoIP망 등 통신망 종류 (가능한 값 : PC, PDA, Telephone, CellularPhone, VoIP, Microphone, BroadcastNews, Other)
레코딩	RecordingEnvironment	녹음 시 발화환경
	RecordingDevice	녹음 시 사용된 마이크, 헤드셋 등 입력장치의 모델명과 주요규격
	TalkingDistance	원거리 음성입력의 경우, 화자와 입력장치와의 발성거리
디지타이징	DigitizingDevice	입력장치로부터 녹취된 음성을 디지타이징하기 위해 사용된 A/D장비의 모델명과 주요규격
	SignalEncoding	디지털 음성데이터의 인코딩 방법
	ByteOrder	상위 바이트와 하위 바이트의 순서
	BitsPerSampling	음성 샘플 당 차지하는 비트 수
	SamplingRate	음성데이터의 샘플링 주파수

4.4. 발화목록 (UtteranceList 엘리먼트)

하나의 발화목록은 발화문장의 집합이다. 발화목록의 식별자는 *Utterancelist_Id* 이다. 발화목록에 대한 설명(UtteranceListDesc)과 발화목록을 구성하는 발화문장의 수만큼 각 발화문장의 정보를 나타내는 Utterance 엘리먼트로 구성된다. 하나의 발화문장은 *Utterance_Id*로 식별한다. <표 4>는 발화목록 정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

4.5. 음성파일 (File 엘리먼트) 정보

음성 DB에 존재하는 각 음성파일에 대한 정보를 기술하는 엘리먼트이다. 해당 음성파일이 누가(*Speaker_Id*) 어떤 문장(*Utterancelist_Id* 및 *Utterance_Id*)을 발화한 것이며, 어떤 채널(*Channel_Id*)로 디지타이징한 것인지를 기술한다. 음성파일 자체 정보로는 식별자(*File_Id*)와 저장위치(FileLocation) 등의 물리적 정보와, 음성구간을 나타내는 *SpeechSection_StartTime*과 *SpeechSection_EndTime* 애트리뷰트, 그리고 음성파일에 존재할 수 있는 다수의 잡음구간을 표시하기 위하여 Noise 엘리먼트를 가진다. 그리고 해당 음성파일에 대한 전사정보를 위한 다수의 Transcription 엘리먼트를 가질 수 있다. <표 5>는 파일정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

<표 3> 화자 엘리먼트 및 속성

분류	엘리먼트 및 속성	내용
화자 기본정보	*Speaker_Id*	화자의 식별자. 화자명의 영문 이니셜 3자리와 숫자 2자리 조합으로 표현. 영문 이니셜이 중복될 경우, 2자리 숫자로 구분
	SpeakerName	발성화자의 한글 이름
	Gender	발성화자의 성별
	Age	발성화자의 나이
	Birthplace	발성화자의 태어난 지역
	EducationPlace	화자가 초등학교 재학 시 거주했던 지역명
	CurrentlyResidenceDuration	발성화자의 현 거주 지역에서의 거주기간
	Job	화자의 직업
	Dialect	화자가 사용하는 방언
	Height	화자의 신장
	Weight	화자의 몸무게
참조	*Utterancelists*	해당화자가 발화한 발화목록 식별자(Utterancelist_Id)의 집합참조
화자 특수정보	*Origin*	외국인인지 내국인인지 표기
	TrainedSpeaker	훈련된 발화자 여부
	SmokingHabit	화자의 흡연 습관
	SpeakingHearingImpairments	화자가 듣거나 말하는 데 장애여부

<표 4> 발화목록 엘리먼트 및 속성

분류	엘리먼트 및 속성	내용
발화목록	*Utterancelist_Id*	발화목록 식별자
	UtteranceListDesc	발화목록 설명
발화문장	Utterance	발화 단어, 문장의 실제내용
	Utterance_Id	발화문장 식별자

4.6. 전사 (Transcription 엘리먼트) 정보

전사정보는 녹음된 파일 하나에 대하여 전사의 종류에 따라 여러 개가 존재할 수 있다. 하나의 음성파일에 해당하는 여러 개의 전사는 *Transcription_Id*로 구분한다. *TranscriptionType*은 Orthographic(철자전사 표기), Morphological(형태소전사 표기), Phonetic(발음전사 표기), Syntactic(구문전사 표기), Semantic(의미전사 표기), Prosodic(운율전사 표기)와 같은 전사 종류를 의미한다. 각 종류별로 해당되는 전사 내용을 Transcription 엘리먼트의 내용으로 기술한다. *Transcriber*는 전사자 정보이다. <표 6>은 전사정보의 주요 엘리먼트 및 속성을 간략하게 소개하고 있다.

<표 5> 파일정보 엘리먼트 및 속성

분류	엘리먼트 및 속성		내용
파일 기본정보	*File_id*		음성파일 식별자
	SpeakingPeriod		녹음 주기와 관련하여 파일별로 발화시점 정보
물리적 정보	FileLocation		음성파일명과 위치
	FileFormat		파일의 포맷, 헤더크기와 파일크기
	SpeechSection_StartTime		음성이 시작하는 시간
	SpeechSection_EndtTime		음성이 끝나는 시간
잡음정보	Noise	NoiseSection	잡음이 시작하는 시간과 끝나는 시간
		NumberOfNoise	음성파일 내 다양한 잡음이 포함되어 있을 경우 잡음의 개수, 잡음개수의 수만큼 잡음구간존재
전사	Transcription		음성파일에 대한 전사정보
참조	*Speaker_id*		해당 음성 파일의 발화자의 식별자 참조
	Channel_id		수집채널의 식별자 참조
	Utterancelist_id		발화목록 식별자 참조
	Utterance_id		발화문장 식별자 참조

<표 6> 전사정보 엘리먼트 및 속성

분류	엘리먼트 및 속성	내용
전사정보	*Transcription_Id*	전사정보 식별자
	TranscriptionType	전사 방법의 종류
	Transcriber	전사정보의 전사기관 또는 전사자의 정보

4.7. 구현 예

<그림 5>는 제안하는 산업용 음성 DB 메타데이타 규격을 특정 음성 DB에 적용한 예이다.

```
<SpeechDBMetadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <General SpeechSetting="Monologue" SpeechStyle="Elicited" Applications="VoiceControl"
   AnnotationStandard="NIST-LDC" DataCategory="VCVSequence">
     <Language>KOR</Language>
     <FullNameOfDataCollection>멀티모달 음성인터페이스를 위한 DB</FullNameOfDataCollection>
     <Project>MVD1</Project>
      <VersionInfo>
                    <Version>1.0</Version>
                    <RecordingPeriod>1999.05.02~2004.11.08</RecordingPeriod>
     </VersionInfo>
     <ReleaseDate>2004-08-13</ReleaseDate>
     <RevisionHistory>
                    <RevisionDate>2004-09-13</RevisionDate>
                    <RevisionDetail>발화자 mkj00의 avi화일 오류 수정</RevisionDetail>
      </RevisionHistory>
     <Copyright>
                    <Authors>홍기형</Authors>
                    <Distributor>Sitec</Distributor>
                    <Performer>MIPS</Performer>
     </Copyright>
     <NumberOfRecordingChannel>3</NumberOfRecordingChannel>
     <Comment>부가내용</Comment>
     <DataVolume>
                    <DataSize>5.0GB</DataSize>
```

```
                <DataDuration>14:20:0Z</DataDuration>
        </DataVolume>
        <Quality OverAllQuality="VeryGood">
                <SoundQualityMeasures>
                        <SNR>8.9dB</SNR>
                        <CrossTalk>4.5dB</CrossTalk>
                        <ClippingRate>64.0%</ClippingRate>
                        <BackgroundNoise>0.03dB</BackgroundNoise>
                        <ErrorRate>0.4%</ErrorRate>
                        <Other>misc</Other>
                </SoundQualityMeasures>
        </Quality>
        <Distribution DistributionCategory="Train" DistributionMedia="VHS" Compression="None"/>
        ....
</General>
<Channels>
 <Channel TypeOfResource="Microphone" Channel_Id="ID01" DataFormat="FloatingPoint">
                <Recording>
                        <RecordingEnvironment>방음실</RecordingEnvironment>
                        <RecordingDevice>
                                <Manufacturer>젠하이져</Manufacturer>
                                <Model>M30</Model>
                                <Spec>dynamic</Spec>
                        </RecordingDevice>
                        <TalkingDistance>35m</TalkingDistance>
                </Recording>
                <Digitizing>
                        <DigitizingDevice>
                                ...
                        </DigitizingDevice>
                        <SignalEncoding type="PCM"/>
                        <ByteOrder order="LitteleEndian"/>
                        <BitsPerSample>8Bit</BitsPerSample>
                        <SamplingRate>150kHZ</SamplingRate>
                </Digitizing>
  </Channel>
  <Channel/>
        ...
</Channels>
<Speakers>
 <Speaker Speaker_Id="csy00" Origin="Native" SmokingHabit="no" TrainedSpeaker="no" Age="20"
  Utterancelists="ID03 ID08" SpeakingHearingImpairment="no" Gender="Female" >
        <SpeakerName>전신영</SpeakerName>
        <BirthPlace>서울</BirthPlace>
        <CurrentlyResidence>서울</CurrentlyResidence>
        <CurrentlyResidenceDuration>20</CurrentlyResidenceDuration>
    ...
  </Speaker>
  <Speaker/>
        ...
</Speakers>
<UtteranceLists>
   <UtteranceList Utterancelist_Id="ID03">
        <UtteranceListDesc>1번부터 105번까지 발화목록</UtteranceListDesc>
        <Utterance  Utterance_Id="ID000001">영</Utterance>
        <Utterance  Utterance_Id="ID000002">공</Utterance>
        ....
        <Utterance/>
  </UtteranceList>
  <UtteranceList/>
  ...
  <UtteranceList/>
</UtteranceLists>
<Files>
   <File  File_Id="ID001" Utterancelist_Id="ID03" Channel_Id="ID01" Utterance_Id="ID000001"
     Speaker_Id ="cjs00">
        <FileFormat format="Wave">
                <HeaderSize>0Byte</HeaderSize>
                <FileSize>1024Byte</FileSize>
        </FileFormat>
        <FileLocation>
                <SpeechFileName>dig01.wav</SpeechFileName>
```

```
                <DirectoryPath>D:₩data₩digit₩female₩f2cjs0</DirectoryPath>
          </FileLocation>
          <Noise NumberOfNoise="1">
                <NoiseSection>
                      <NoiseType>음악</NoiseType>
                      <NoiseStartTime>00:01:34</NoiseStartTime>
                      <NoiseEndTime>00:01:50</NoiseEndTime>
                </NoiseSection>
          </Noise>
          <Transcription TranscriptionType="Syntactic" Transcription_Id="ID0001"
          Transcriber="Sitec">영</Transcription>
    </File>
    <File/>
     ...
 </Files>
</SpeechDBMetadata>
```

<그림 5> XML 기반 음성 DB 메타데이터 예제

5. 결론 및 향후과제

본 논문에서는 산업용 음성 DB를 위한 메타데이터 규격을 XML 기반으로 제안하였다. 본 규격은 ETRI와 SiTEC에서 기존에 제안한 규격[5]를 객체 지향 구조로 보완하였다. 화자, 채널, 발화목록, 파일, 전사 정보에 각각의 식별자를 두고 구성요소 사이의 관계를 식별자의 참조로 표시할 수 있도록 하였다. 또한 연관되는 엘리먼트들의 그룹핑과 계층구조를 도입하여, 메타데이터의 구조를 체계화하여, 향후 음성 DB의 검색이나 재사용이 용이하도록 설계하였다.

향후 계획으로는 대용량의 산업용 음성 DB를 본 규격에 따라 쉽게 메타데이터를 구축할 수 있도록 하는 산업용 음성 DB 메타데이터 구축도구의 개발을 진행하려고 한다. 또한, 구축한 메타데이터로부터 사용자가 필요한 정보를 검색할 수 있는 도구를 개발하여, 배포할 예정이다.

참 고 문 헌

[1] ELRA http://www.elra.info/services/speech_1.4.rtf, 2004.

[2] LDC http://www.ldc.upenn.edu/Catalog, 2004.

[3] 홍기형, 이욱재, "국제 음성기술 표준화 동향과 대응", *음성통신 및 신호처리 학술대회 논문집,* 20권, 1호, pp.185-188, 2003.

[4] 김상훈, 이용주, "음성 DB 표준화", *음성통신 및 신호처리 학술대회 논문집,* 20권, 1호, pp.181-184, 2003.

[5] 김상훈, 이영직, 한민수, "음성 DB 부가 정보 기술방안 표준화를 위한 제안", *말소리,* 제47호, pp.110-119, 2003.

[6] XML Schema http://www.w3.org/TR/xmlschema-0/, 2004.

접수일자 : 2005년 8월 19일
게재결정 : 2005년 9월 13일

▶ 주영희(Young-Hee Joo)
주소: 136-742 서울시 동선동 3가 249-1 성신여자대학교
소속: 성신여자대학교 MIPS 연구실
전화: 02) 928-9997
E-mail: yhjoo@media.sungshin.ac.kr

▶ 홍기형(Ki-Hyung Hong)
주소: 136-742 서울시 동선동 3가 249-1 성신여자대학교
소속: 성신여자대학교 미디어정보학부 MIPS 연구실
전화: 02) 920-7525
E-mail: khhong@sungshin.ac.kr